

Best Available Copy



⑯ Europäisches Patentamt
European Patent Office
Office européen des brevets



⑪ Publication number: 0 422 697 B1

⑫ EUROPEAN PATENT SPECIFICATION

⑯ Date of publication of patent specification: 30.03.94 ⑯ Int. Cl. 5: C12N 15/21, C07K 13/00

⑯ Application number: 90124236.2

⑯ Date of filing: 25.04.83

⑯ Publication number of the earlier application in
accordance with Art.76 EPC: 0 108 128

⑯ A consensus leukocyte interferon and a gene encoding it.

⑯ Priority: 06.05.82 US 375494
15.04.83 US 483451

⑯ Date of publication of application:
17.04.91 Bulletin 91/16

⑯ Publication of the grant of the patent:
30.03.94 Bulletin 94/13

⑯ Designated Contracting States:
AT BE CH DE FR GB LI LU NL SE

⑯ References cited:

THE BIOLOGY OF THE INTERFERON SYSTEM,
PROCEEDINGS, 2ND INTERNATIONAL
TNOMEETING, 18th - 22nd April 1983, pages
119-127, Elsevier Science Publishers B.V.; K.
ALTON et al.: "Production, characterization
and biological effects of recombinant DNA
derived human IFN-alpha and IFN-gamma
analogues"

NATURE, vol. 292, 20th August 1981, pages
756-762, Macmillan Journals Ltd; M.D.EDGE
et al.: "Total synthesis of a human leukocyte
interferon gene"

⑯ Proprietor: AMGEN INC.
1840 Dehavilland Drive
Thousand Oaks California 91320 -1789(US)

⑯ Inventor: Alton, Norman K.
815 Country Valley Road
Thousand Oaks, California 91362(US)
Inventor: Peters, Mary A.
945,37th Street
Boulder, CO 80303(US)
Inventor: Stabinsky, Yitzhak
Peprotech Inc./5 Crescent Avenue
Rockyhill, New Jersey 08553(US)
Inventor: Snitman, David L.
1475 Ithica Drive
Boulder, CO 80303(US)

⑯ Representative: Brown, John David et al
FORRESTER & BOEHMERT
Franz-Joseph-Strasse 38
D-80801 München (DE)

EP 0 422 697 B1

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

EP 0 422 697 B1

NATURE, vol. 290, no. 5801, 5th March 1981,
pages 20-26, Macmillan Journals

Ltd, Chesham, Bucks, GB; D.V. GOEDDEL et
al.: "The structure of eight distinct cloned
human leukocyte interferon cDNAs"

Description

The present invention relates generally to the manipulation of genetic materials and, more particularly, to the manufacture of specific DNA sequences useful in recombinant procedures to secure the production of proteins of interest.

5 Genetic materials may be broadly defined as those chemical substances which program for and guide the manufacture of constituents of cells and viruses and direct the responses of cells and viruses. A long chain polymeric substance known as deoxyribonucleic acid (DNA) comprises the genetic material of all living cells and viruses except for certain viruses which are programmed by ribonucleic acids (RNA). The 10 repeating units in DNA polymers are four different nucleotides, each of which consists of either a purine (adenine or guanine) or a pyrimidine (thymine or cytosine) bound to a deoxyribose sugar to which a phosphate group is attached. Attachment of nucleotides in linear polymeric form is by means of fusion of the 5' phosphate of one nucleotide to the 3' hydroxyl group of another. Functional DNA occurs in the form of stable double stranded associations of single strands of nucleotides (known as deoxyoligonucleotides), 15 which associations occur by means of hydrogen bonding between purine and pyrimidine bases [i.e., "complementary" associations existing either between adenine (A) and thymine (T) or guanine (G) and cytosine (C)]. By convention, nucleotides are referred to by the names of their constituent purine or pyrimidine bases, and the complementary associations of nucleotides in double stranded DNA (i.e., A-T and G-C) are referred to as "base pairs". Ribonucleic acid is a polynucleotide comprising adenine, guanine, 20 cytosine and uracil (U), rather than thymine, bound to ribose and a phosphate group.

Most briefly put, the programming function of DNA is generally effected through a process wherein specific DNA nucleotide sequences (genes) are "transcribed" into relatively unstable messenger RNA (mRNA) polymers. The mRNA, in turn, serves as a template for the formation of structural, regulatory and catalytic proteins from amino acids. This translation process involves the operations of small RNA strands 25 (tRNA) which transport and align individual amino acids along the mRNA strand to allow for formation of polypeptides in proper amino acid sequences. The mRNA "message", derived from DNA and providing the basis for the tRNA supply and orientation of any given one of the twenty amino acids for polypeptide "expression", is in the form of triplet "codons" -- sequential groupings of three nucleotide bases. In one sense, the formation of a protein is the ultimate form of "expression" of the programmed genetic message 30 provided by the nucleotide sequence of a gene.

Certain DNA sequences which usually "precede" a gene in a DNA polymer provide a site for initiation of the transcription into mRNA. These are referred to as "promoter" sequences. Other DNA sequences, also usually "upstream" of (i.e., preceding) a gene in a given DNA polymer, bind proteins that determine the frequency (or rate) of transcription initiation. These other sequences are referred to as "regulator" 35 sequences. Thus, sequences which precede a selected gene (or series of genes) in a functional DNA polymer and which operate to determine whether the transcription (and eventual expression) of a gene will take place are collectively referred to as "promoter/regulator" or "control" DNA sequences. DNA sequences which "follow" a gene in a DNA polymer and provide a signal for termination of the transcription into mRNA are referred to as "terminator" sequences.

40 A focus of microbiological processing for nearly the last decade has been the attempt to manufacture industrially and pharmaceutically significant substances using organisms which do not initially have genetically coded information concerning the desired product included in their DNA. Simply put, a gene that specifies the structure of a product is either isolated from a "donor" organism or chemically synthesized and then stably introduced into another organism which is preferably a self-replicating unicellular microorganism. Once this is done, the existing machinery for gene expression in the "transformed" host cells 45 operates to construct the desired product.

The art is rich in patent and literature publications relating to "recombinant DNA" methodologies for the isolation, synthesis, purification and amplification of genetic materials for use in the transformation of selected host organisms. U.S. Letters Patent No. 4,237,224 to Cohen, et al., for example, relates to 50 transformation of prokaryotic unicellular host organisms with "hybrid" viral or circular plasmid DNA which includes selected exogenous DNA sequences. The procedures of the Cohen, et al. patent first involve manufacture of a transformation vector by enzymatically cleaving viral or circular plasmid DNA to form linear DNA strands. Selected foreign DNA strands are also prepared in linear form through use of similar enzymes. The linear viral or plasmid DNA is incubated with the foreign DNA in the presence of ligating 55 enzymes capable of effecting a restoration process and "hybrid" vectors are formed which include the selected foreign DNA segment "spliced" into the viral or circular DNA plasmid.

Transformation of compatible unicellular host organisms with the hybrid vector results in the formation of multiple copies of the foreign DNA in the host cell population. In some instances, the desired result is

simply the amplification of the foreign DNA and the "product" harvested is DNA. More frequently, the goal of transformation is the expression by the host cells of the foreign DNA in the form of large scale synthesis of isolatable quantities of commercially significant protein or polypeptide fragments coded for by the foreign DNA. See also, e.g., U.S. Letters Patent Nos. 4,269,731 (to Shine), 4,273,875 (to Manis) and 4,293,652 (to Cohen).

The success of procedures such as described in the Cohen, et al. patent is due in large part to the ready availability of "restriction endonuclease" enzymes which facilitate the site-specific cleavage of both the unhybridized DNA vector and, e.g., eukaryotic DNA strands containing the foreign sequences of interest. Cleavage in a manner providing for the formation of single stranded complementary "ends" on the double stranded linear DNA strands greatly enhances the likelihood of functional incorporation of the foreign DNA into the vector upon "ligating" enzyme treatment. A large number of such restriction endonuclease enzymes are currently commercially available [See, e.g., "BRL Restriction Endonuclease Reference Chart" appearing in the "81/82 Catalog" of Bethesda Research Laboratories, Inc., Gaithersburg, Maryland.] Verification of hybrid formation is facilitated by chromatographic techniques which can, for example, distinguish the hybrid plasmids from non-hybrids on the basis of molecular weight. Other useful verification techniques involve radioactive DNA hybridization.

Another manipulative "tool" largely responsible for successes in transformation of prokaryotic cells is the use of selectable "marker" gene sequences. Briefly put, hybrid vectors are employed which contain, in addition to the desired foreign DNA, one or more DNA sequences which code for expression of a phenotypic trait capable of distinguishing transformed from non-transformed host cells. Typical marker gene sequences are those which allow a transformed prokaryotic cell to survive and propagate in a culture medium containing metals, antibiotics, and like components which would kill or severely inhibit propagation of non-transformed host cells.

Successful expression of an exogenous gene in a transformed host microorganism depends to a great extent on incorporation of the gene into a transformation vector with a suitable promoter/regulator region present to insure transcription of the gene into mRNA and other signals which insure translation of the mRNA message into protein (e.g., ribosome binding sites). It is not often the case that the "original" promoter/regulator region of a gene will allow for high levels of expression in the new host. Consequently, the gene to be inserted must either be fitted with a new, host-accommodated transcription and translation regulating DNA sequence prior to insertion or it must be inserted at a site where it will come under the control of existing transcription and translation signals in the vector DNA.

It is frequently the case that the insertion of an exogenous gene into, e.g., a circular DNA plasmid vector, is performed at a site either immediately following an extant transcription and translation signal or within an existing plasmid-borne gene coding for a rather large protein which is the subject of high degrees of expression in the host. In the latter case, the host's expression of the "fusion gene" so formed results in high levels of production of a "fusion protein" including the desired protein sequence (e.g., as an intermediate segment which can be isolated by chemical cleavage of large protein). Such procedures not only insure desired regulation and high levels of expression of the exogenous gene product but also result in a degree of protection of the desired protein product from attack by proteases endogenous to the host. Further, depending on the host organism, such procedures may allow for a kind of "piggyback" transportation of the desired protein from the host cells into the cell culture medium, eliminating the need to destroy host cells for the purpose of isolating the desired product.

While the foregoing generalized descriptions of published recombinant DNA methodologies may make the processes appear to be rather straightforward, easily performed and readily verified, it is actually the case that the DNA sequence manipulations involved are quite painstakingly difficult to perform and almost invariably characterized by very low yields of desired products.

As an example, the initial "preparation" of a gene for insertion into a vector to be used in transformation of a host microorganism can be an enormously difficult process, especially where the gene to be expressed is endogenous to a higher organism such as man. One laborious procedure practiced in the art is the systematic cloning into recombinant plasmids of the total DNA genome of the "donor" cells, generating immense "libraries" of transformed cells carrying random DNA sequence fragments which must be individually tested for expression of a product of interest. According to another procedure, total mRNA is isolated from high expression donor cells (presumptively containing multiple copies of mRNA coded for the product of interest), first "copied" into single stranded cDNA with reverse transcriptase enzymes, then into double stranded form with polymerase, and cloned. The procedure again generates a library of transformed cells somewhat smaller than a total genome library which may include the desired gene copies free of non-transcribed "introns" which can significantly interfere with expression by a host microorganism. The above-noted time-consuming gene isolation procedures were in fact employed in published recombinant DNA

procedures for obtaining microorganism expression of several proteins, including rat proinsulin [Ullrich, et al., Science, 196, pp. 1313-1318 (1977)], human fibroblast interferon [Goedell, et al., Nucleic Acids Research, 8, pp. 4087-4094 (1980)], mouse β -endorphin [Shine, et al., Nature, 285, pp. 456-461 (1980)] and human leukocyte interferon [Goedell, et al., Nature, 287, pp. 411-416 (1980); and Goedell, et al., Nature, 290, pp. 20-26 (1981)].

Whenever possible, the partial or total manufacture of genes of interest from nucleotide bases constitutes a much preferred procedure for preparation of genes to be used in recombinant DNA methods. A requirement for such manufacture is, of course, knowledge of the correct amino acid sequence of the desired polypeptide. With this information in hand, a generative DNA sequence code for the protein (i.e., a properly ordered series of base triplet codons) can be planned and a corresponding synthetic, double stranded DNA segment can be constructed. A combination of manufacturing and cDNA synthetic methodologies is reported to have been employed in the generation of a gene for human growth hormone. Specifically, a manufactured linear double stranded DNA sequence of 72 nucleotide base pairs (comprising codons specifying the first 24 amino acids of the desired 191 amino acid polypeptide) was ligated to a cDNA-derived double strand coding for amino acids Nos. 25-191 and inserted in a modified pBR322 plasmid at a locus controlled by a lac promotor/regulator sequence [Goedell, et al., Nature, 281, pp. 544-548 (1981)].

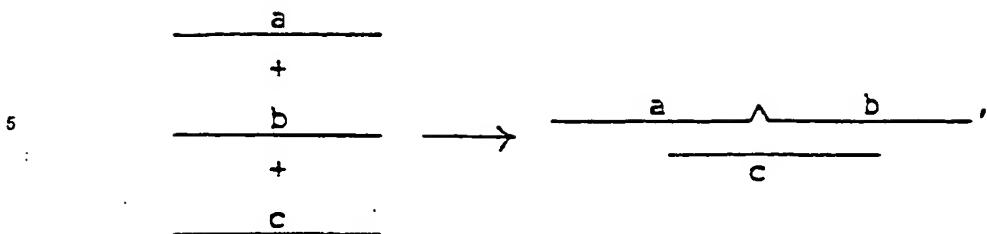
Completely synthetic procedures have been employed for the manufacture of genes coding for relatively "short" biologically functional polypeptides, such as human somatostatin (14 amino acids) and human insulin (2 polypeptide chains of 21 and 30 amino acids, respectively).

In the somatostatin gene preparative procedure [Itakura, et al., Science, 198, pp. 1056-1063 (1977)] a 52 base pair gene was constructed wherein 42 base pairs represented the codons specifying the required 14 amino acids and an additional 10 base pairs were added to permit formation of "sticky-end" single stranded terminal regions employed for ligating the structural gene into a microorganism transformation vector. Specifically, the gene was inserted close to the end of a β -galactosidase enzyme gene and the resultant fusion gene was expressed as a fusion protein from which somatostatin was isolated by cyanogen bromide cleavage. Manufacture of the human insulin gene, as noted above, involved preparation of genes coding for a 21 amino acid chain and for a 30 amino acid chain. Eighteen deoxyoligonucleotide fragments were combined to make the gene for the longer chain, and eleven fragments were joined into a gene for the shorter chain. Each gene was employed to form a fusion gene with a β -galactosidase gene and the individually expressed polypeptide chains were enzymatically isolated and linked to form complete insulin molecules. [Goedell, et al., Proc. Nat. Acad. Sci. U.S.A., 76, pp. 106-110 (1979).]

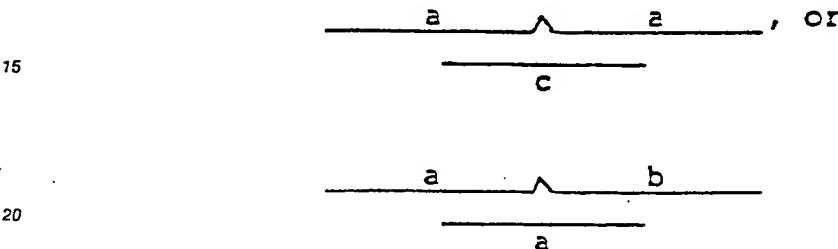
In each of the above procedures, deoxyoligonucleotide segments were prepared, and then sequentially ligated according to the following general procedure. [See, e.g., Agarwal, et al., Nature, 227, pp. 1-7 (1970) and Khorana, Science, 203, pp. 614-675 (1979)]. An initial "top" (i.e., 5'-3' polarity) deoxyoligonucleotide segment is enzymatically joined to a second "top" segment. Alignment of these two "top" strands is made possible using a "bottom" (i.e., 3' to 5' polarity) strand having a base sequence complementary to half of the first top strand and half of the second top strand. After joining, the uncomplemented bases of the top strands "protrude" from the duplex portion formed. A second bottom strand is added which includes the five or six base complement of a protruding top strand, plus an additional five or six bases which then protrude as a bottom single stranded portion. The two bottom strands are then joined. Such sequential additions are continued until a complete gene sequence is developed, with the total procedure being very time-consuming and highly inefficient.

The time-consuming characteristics of such methods for total gene synthesis are exemplified by reports that three months' work by at least four investigators was needed to perform the assembly of the two "short", insulin genes previously referred to. Further, while only relatively small quantities of any manufactured gene are needed for success of vector insertion, the above synthetic procedures have such poor overall yields (on the order of 20% per ligation) that the eventual isolation of even minute quantities of a selected short gene is by no means guaranteed with even the most scrupulous adherence to prescribed methods. The maximum length gene which can be synthesized is clearly limited by the efficiency with which the individual short segments can be joined. If n such ligation reactions are required and the yield of each such reaction is y , the quantity of correctly synthesized genetic material obtained will be proportional to y^n . Since this relationship is exponential in nature, even a small increase in the yield per ligation reaction will result in a substantial increase in the length of the largest gene that may be synthesized.

Inefficiencies in the above-noted methodology are due in large part to the formation of undesired intermediate products. As an example, in an initial reaction forming annealed top strands associated with a bottom, "template" strand, the desired reaction may be,



but the actual products obtained may be



25 or the like. Further, the longer the individual deoxyoligonucleotides are, the more likely it is that they will form thermodynamically stable self-associations such as "hairpins" or aggregations.

Proposals for increasing synthetic efficiency have not been forthcoming and it was recently reported that, "With the methods now available, however, it is not economically practical to synthesize genes for peptides longer than about 30 amino acid units, and many clinically important proteins are much longer". [Aharonowitz, et al. Scientific American, 245, No. 3, pp. 140-152, at p. 151 (1981).]

30 An illustration of the "economic practicalities" involved in large gene synthesis is provided by the recent publication of "successful" efforts in the total synthesis of a human leukocyte interferon gene. [Edge, et al., Nature, 292, pp. 756-782 (1981).] Briefly summarized, 67 different deoxyoligonucleotides containing about 15 bases were synthesized and joined in the "50 percent overlap" procedure of the type noted above to form eleven short duplexes. These, in turn were assembled into four longer duplexes which were 35 eventually joined to provide a 514 base pair gene coding for the 166 amino acid protein. The procedure, which the authors characterize as "rapid", is reliably estimated to have consumed nearly a year's effort by five workers and the efficiency of the assembly strategy was clearly quite poor. It may be noted, for example, that while 40 pmole of each of the starting 67 deoxyoligonucleotides was prepared and employed to form the eleven intermediate-sized duplexes, by the time assembly of the four large duplexes was 40 achieved, a yield of only about 0.01 pmole of the longer duplexes could be obtained for use in final assembly of the whole gene.

Another aspect of the practice of recombinant DNA techniques for the expression, by microorganisms, of proteins of industrial and pharmaceutical interest is the phenomenon of "codon preference". While it was earlier noted that the existing machinery for gene expression in genetically transformed host cells will 45 "operate" to construct a given desired product, levels of expression attained in a microorganism can be subject to wide variation, depending in part on specific alternative forms of the amino acid-specifying genetic code present in an inserted exogenous gene. A "triplet" codon of four possible nucleotide bases can exist in 64 variant forms. That these forms provide the message for only 20 different amino acids (as well as transcription initiation and termination) means that some amino acids can be coded for by more than 50 one codon. Indeed, some amino acids have as many as six "redundant", alternative codons while some others have a single, required codon. For reasons not completely understood, alternative codons are not at all uniformly present in the endogenous DNA of differing types of cells and there appears to exist a variable natural hierarchy or "preference" for certain codons in certain types of cells.

As one example, the amino acid leucine is specified by any of six DNA codons including CTA, CTC, 55 CTG, CTT, TTA, and TTG (which correspond, respectively, to the mRNA codons, CUA, CUC, CUG, CUU, UUA and UUG). Exhaustive analysis of genome codon frequencies for microorganisms has revealed endogenous DNA of E. coli bacteria most commonly contains the CTG leucine-specifying codon, while the DNA of yeasts and slime molds most commonly includes a TTA leucine-specifying codon. In view of this

hierarchy, it is generally held that the likelihood of obtaining high levels of expression of a leucine-rich polypeptide by an E. coli host will depend to some extent on the frequency of codon use. For example, a gene rich in TTA codons will in all probability be poorly expressed in E. coli, whereas a CTG rich gene will probably highly express the polypeptide. In a like manner, when yeast cells are the projected transformation host cells for expression of a leucine-rich polypeptide, a preferred codon for use in an inserted DNA would be TTA. See, e.g., Grantham, et al., Nucleic Acids Research, 8, pp. r49-r62 (1980); Grantham, et al., Nucleic Acids Research, 8, pp. 1893-1912 (1980); and, Grantham, et al., Nucleic Acids Research, 9, pp. r43-r74 (1981).

The implications of codon preference phenomena on recombinant DNA techniques are manifest, and the phenomenon may serve to explain many prior failures to achieve high expression levels for exogenous genes in successfully transformed host organisms — a less "preferred" codon may be repeatedly present in the inserted gene and the host cell machinery for expression may not operate as efficiently. This phenomenon directs the conclusion that wholly manufactured genes which have been designed to include a projected host cell's preferred codons provide a preferred form of foreign genetic material for practice of recombinant DNA techniques. In this context, the absence of procedures for rapid and efficient total gene manufacture which would permit codon selection is seen to constitute an even more serious roadblock to advances in the art.

Of substantial interest to the background of the present invention is the state of the art with regard to the preparation and use of a class of biologically active substances, the interferons (IFNs). Interferons are secreted proteins having fairly well-defined antiviral, antitumor and immunomodulatory characteristics. See, e.g., Gray, et al., Nature, 295, pp. 503-508 (1982) and Edge, et al., supra, and references cited therein.

On the basis of antigenicity and biological and chemical properties, human interferons have been grouped into three major classes: IFN- α (leukocyte), IFN- β (fibroblast) and IFN- γ (immune). Considerable information has accumulated on the structures and properties of the virus-induced acid-stable interferons (IFN- α and β) (leukocyte interferon has been referred to as "LeIFN" and "IFN- α " and is also generally called " α -IFN"). These have been purified to homogeneity and at least partial amino acid sequences have been determined. Analyses of cloned cDNA and gene sequences for IFN- β_1 and the IFN- α multigene family have permitted the deduction of the complete amino acid sequences of many of the interferons. In addition, efficient synthesis of IFN- β_1 and several IFN- α s in E. coli, and IFN- α_1 , in yeast, have now made possible the purification of large quantities of these proteins in biologically active form.

Much less information is available concerning the structure and properties of IFN- γ , an interferon generally produced in cultures of lymphocytes exposed to various mitogenic stimuli. It is acid labile and does not cross-react with antisera prepared against IFN- α or IFN- β . A broad range of biological activities have been attributed to IFN- γ including potentiation of the antiviral activities of IFN- α and - β , from which it differs in terms of its virus and cell specificities and the antiviral mechanisms induced.

The above-noted wide variations in biological activities of various interferon types makes the construction of synthetic polypeptide analogs of the interferons of paramount significance to the full development of the therapeutic potential of this class of compounds. Despite the advantages in isolation of quantities of interferons which have been provided by recombinant DNA techniques to date, practitioners in this field have not been able to address the matter of preparation of synthetic polypeptide analogs of the interferons with any significant degree of success.

Put another way, the work of Gray, et al., supra, in the isolation of a gene coding for IFN- γ and the extensive labors of Edge, et al., supra, in providing a wholly manufactured IFN- α_1 gene provide only genetic materials for expression of single, very precisely defined, polypeptide sequences. There exist no procedures (except, possibly, for site specific mutagenesis) which would permit microbial expression of large quantities of human IFN- γ analogs which differed from the "authentic" polypeptide in terms of the identity or location of even a single amino acid. In a like manner, preparation of an IFN- α_1 analog which differed by one amino acid from the polypeptide prepared by Edge, et al., supra, would appear to require an additional year of labor in constructing a whole new gene which varied in terms of a single triplet codon. No means is readily available for the excision of a fragment of the subject gene and replacement with a fragment including the coding information for a variant polypeptide sequence. Further, modification of the reported cDNA-derived and manufactured DNA sequences to vary codon usage is not an available "option".

Indeed, the only report of the preparation of variant interferon polypeptide species by recombinant DNA techniques has been in the context of preparation and expression of "hybrids" of human genes for IFN- α_1 and IFN- α_2 [Weck, et al., Nucleic Acids Research, 9, pp. 6153-6168 (1981) and Streuli, et al., Proc. Nat. Acad. Sci. U.S.A., 78, pp. 2848-2852 (1981)]. The hybrids obtained consisted of the four possible combinations of gene fragments developed upon finding that two of the eight human (cDNA-derived) genes fortuitously included only once within the sequence, base sequences corresponding to the restriction

endonuclease cleavage sites for the bacterial endonucleases, Pvull and BgIII.

There exists, therefore, a substantial need in the art for more efficient procedures for the total synthesis from nucleotide bases of manufactured DNA sequences coding for large polypeptides such as the interferons. There additionally exists a need for synthetic methods which will allow for the rapid construction 5 of variant forms of synthetic sequences such as will permit the microbial expression of synthetic polypeptides which vary from naturally occurring forms in terms of the identity and/or position of one or more selected amino acids.

BRIEF SUMMARY

10 The present invention provides novel, rapid and highly efficient procedures for the total synthesis of linear, double stranded DNA sequences in excess of about 200 nucleotide base pairs in length, which sequences may comprise entire structural genes capable of directing the synthesis of a wide variety of polypeptides of interest.

15 According to the invention, linear, double stranded DNA sequences of a length in excess of about 200 base pairs and coding for expression of a predetermined continuous sequence of amino acids within a selected host microorganism transformed by a selected DNA vector including the sequence, are synthesized by a method comprising:

20 (a) preparing two or more different, subunit, linear, double stranded DNA sequences of about 100 or more base pairs in length for assembly in a selected assembly vector,

each different subunit DNA sequence prepared comprising a series of nucleotide base codons coding for a different continuous portion of said predetermined sequence of amino acids to be expressed,

25 one terminal region of a first of said subunits comprising a portion of a base sequence which provides a recognition site for cleavage by a first restriction endonuclease, which recognition site is entirely present either once or not at all in said selected assembly vector upon insertion of the subunit therein,

30 one terminal region of a second of said subunits comprising a portion of a base sequence which provides a recognition site for cleavage by a second restriction endonuclease other than said first endonuclease, which recognition site is entirely present once or not at all in said selected assembly vector upon insertion of the subunit therein,

35 at least one-half of all remaining terminal regions of subunits comprising a portion of a recognition site (preferably a palindromic six base recognition site) for cleavage by a restriction endonuclease other than said first and second endonucleases, which recognition site is entirely present once and only once in said selected assembly vector after insertion of all subunits thereinto; and

40 (b) serially inserting each of said subunit DNA sequences prepared in step (a) into the selected assembly vector and effecting the biological amplification of the assembly vector subsequent to each insertion, thereby to form a DNA vector including the desired DNA sequence coding for the predetermined continuous amino acid sequence and wherein the desired DNA sequence assembled includes at least one unique, preferably palindromic six base, recognition site for restriction endonuclease cleavage at an intermediate position therein.

45 The above general method preferably further includes the step of isolating the desired DNA sequence from the assembly vector preferably to provide one of the class of novel manufactured DNA sequences having at least one unique palindromic six base recognition site for restriction endonuclease cleavage at an intermediate position therein. A sequence so isolated may then be inserted in a different, "expression" vector and direct expression of the desired polypeptide by a microorganism which is the same as or different from that in which the assembly vector is amplified. In other preferred embodiments of the method: at least three different subunit DNA sequences are prepared in step (a) and serially inserted into said selected assembly vector in step (b) and the desired manufactured DNA sequence obtained includes at 50 least two unique palindromic six base recognition sites for restriction endonuclease cleavage at intermediate positions therein; the DNA sequence synthesized comprises an entire structural gene coding for a biologically active polypeptide; and, in the DNA sequence manufactured, the sequence of nucleotide bases includes one or more codons selected, from among alternative codons specifying the same amino acid, on the basis of preferential expression characteristics of the codon in said selected host microorganism.

55 Novel products of the invention include manufactured, linear, double stranded DNA sequences of a length in excess of about 200 base pairs and coding for the expression of a predetermined continuous sequence of amino acids by a selected host microorganism transformed with a selected DNA vector including the sequence, characterized by having at least one unique palindromic six base recognition site

for restriction endonuclease cleavage at an intermediate position therein. Also included are polypeptide products of the expression by an organism of such manufactured sequences.

Illustratively provided by the present invention are novel manufactured genes coding for consensus human leukocyte interferons.

5 DNA subunit sequences for use in practice of the methods of the invention are preferably synthesized from nucleotide bases according to the methods disclosed in co-owned, concurrently-filed U.S. Patent Application Serial No. 375,493, by Yitzhak Stabinsky, entitled "Manufacture and Expression of Structural Genes" (Attorney's Docket No. 6250). Briefly summarized the general method comprises the steps of:

10 (1) preparing two or more different, linear, duplex DNA strands, each duplex strand including a double stranded region of 12 or more selected complementary base pairs and further including a top single stranded terminal sequence of from 3 to 7 selected bases at one end of the strand and/or a bottom single stranded terminal sequence of from 3 to 7 selected bases at the other end of the strand, each single stranded terminal sequence of each duplex DNA strand comprising the entire base complement of at most one single stranded terminal sequence of any other duplex DNA strand prepared; and

15 (2) annealing each duplex DNA strand prepared in step (1) to one or two different duplex strands prepared in step (1) having a complementary single stranded terminal sequence, thereby to form a single continuous double stranded DNA sequence which has a duplex region of at least 27 selected base pairs including at least 3 base pairs formed by complementary association of single stranded terminal sequences of duplex DNA strands prepared in step (1) and which has from 0 to 2 single stranded top or

20 bottom terminal regions of from 3 to 7 bases.

In the preferred general process for subunit manufacture, at least three different duplex DNA strands are prepared in step (1) and all strands so prepared are annealed concurrently in a single annealing reaction mixture to form a single continuous double stranded DNA sequence which has a duplex region of at least 42 selected base pairs including at least two non-adjacent sets of 3 or more base pairs formed by

25 complementary association of single stranded terminal sequences of duplex strands prepared in step (1).

The duplex DNA strand preparation step (1) of the preferred subunit manufacturing process preferably comprises the steps of:

30 (a) constructing first and second linear deoxyoligonucleotide segments having 15 or more bases in a selected linear sequence, the linear sequence of bases of the second segment comprising the total complement of the sequence of bases of the first segment except that at least one end of the second segment shall either include an additional linear sequence of from 3 to 7 selected bases beyond those fully complementing the first segment, or shall lack a linear sequence of from 3 to 7 bases complementary to a terminal sequence of the first segment, provided, however, that the second segment shall not have an additional sequence of bases or be lacking a sequence of bases at both of its ends; and,

35 (b) combining the first and second segments under conditions conducive to complementary association between segments to form a linear, duplex DNA strand.

The sequence of bases in the double stranded DNA subunit sequences formed preferably includes one or more triplet codons selected from among alternative codons specifying the same amino acid on the basis of preferential expression characteristics of the codon in a projected host microorganism, such as

40 yeast cells or bacteria, especially *E. coli* bacteria.

Also provided by the present invention are improvements in methods and materials for enhancing levels of expression of selected exogenous genes in *E. coli* host cells. Briefly stated, expression vectors are constructed to include selected DNA sequences upstream of polypeptide coding regions which selected sequences are duplicative of ribosome binding site sequences extant in genomic *E. coli* DNA associated

45 with highly expressed endogenous polypeptides. A presently preferred selected sequence is duplicative of the ribosome binding site sequence associated with *E. coli* expression of outer membrane protein F ("OMP-F").

Other aspects and advantages of the present invention will be apparent upon consideration of the following detailed description thereof.

50

DETAILED DESCRIPTION

As employed herein, the term "manufactured" as applied to a DNA sequence or gene shall designate a product either totally chemically synthesized by assembly of nucleotide bases or derived from the biological replication of a product thus chemically synthesized. As such, the term is exclusive of products "synthesized" by cDNA methods or genomic cloning methodologies which involve starting materials which are of biological origin. Table I below sets out abbreviations employed herein to designate amino acids and includes IUPAC-recommended single letter designations.

TABLE I

	<u>Amino Acid</u>	<u>Abbreviation</u>	<u>IUPAC Symbol</u>
	Alanine	Ala	A
10	Cysteine	Cys	C
	Aspartic acid	Asp	D
	Glutamic acid	Glu	E
15	Phenylalanine	Phe	F
	Glycine	Gly	G
	Histidine	His	H
20	Isoleucine	Ile	I
	Lysine	Lys	K
	Leucine	Leu	L
25	Methionine	Met	M
	Asparagine	Asn	N
	Proline	Pro	P
30	Glutamine	Gln	Q
	Arginine	Arg	R
	Serine	Ser	S
35	Threonine	Thr	T
	Valine	Val	V
	Tryptophan	Trp	W
40	Tyrosine	Tyr	Y

The following abbreviations shall be employed for nucleotide bases: A for adenine; G for guanine; T for thymine; U for uracil; and C for cytosine. For ease of understanding of the present invention, Table II and II below provide tabular correlations between the 64 alternate triplet nucleotide base codons of DNA and the 45 20 amino acids and transcription termination ("stop") functions specified thereby. In order to determine the corresponding correlations for RNA, U is substituted for T in the tables.

TABLE II

5	FIRST POSITION	SECOND POSITION			THIRD POSITION	
		T	C	A		
10	T	Phe	Ser	Tyr	Cys	T
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	Stop	Stop	A
		Leu	Ser	Stop	Trp	G
15	C	Leu	Pro	His	Arg	T
		Leu	Pro	His	Arg	C
		Leu	Pro	Gln	Arg	A
		Leu	Pro	Gln	Arg	G
20	A	Ile	Thr	Asn	Ser	T
		Ile	Thr	Asn	Ser	C
		Ile	Thr	Lys	Arg	A
		Met	Thr	Lys	Arg	G
25	G	Val	Ala	Asp	Gly	T
		Val	Ala	Asp	Gly	C
		Val	Ala	Glu	Gly	A
		Val	Ala	Glu	Gly	G
40						

45

50

55

TABLE III

<u>Amino Acid</u>	<u>Specifying Codon(s)</u>
(A) Alanine	GCT, GCC, GCA, GCG
(C) Cysteine	TGT, TGC
(D) Aspartic acid	GAT, GAC
(E) Glutamic acid	GAA, GAG
(F) Phenylalanine	TTT, TTC
(G) Glycine	GGT, GGC, GGA, GGG
(H) Histidine	CAT, CAC
(I) Isoleucine	ATT, ATC, ATA
(K) Lysine	AAA, AAG
(L) Leucine	TTA, TTG, CTT, CTC, CTA, CTG
(M) Methionine	ATG
(N) Asparagine	AAT, AAC
(P) Proline	CCT, CCC, CCA, CCG
(Q) Glutamine	CAA, CAG
(R) Arginine	CGT, CGC, CGA, CGG, AGA, AGG
(S) Serine	TCT, TCC, TCA, TCG, AGT, AGC
(T) Threonine	ACT, ACC, ACA, ACG
(V) Valine	GTT, GTC, GTA, GTG
(W) Tryptophan	TGG
(Y) Tyrosine	TAC, TAT
STOP	TAA, TAG, TGA

A "palindromic" recognition site for restriction endonuclease cleavage of double stranded DNA is one which displays "left-to-right and right-to-left" symmetry between top and bottom base complements, i.e., where "readings" of complementary base sequences of the recognition site from 5' to 3' ends are identical. Examples of palindromic six base recognition sites for restriction endonuclease cleavage include the sites for cleavage by HindIII wherein top and bottom strands read from 5' to 3' as AAGCTT. A non-palindromic six base restriction site is exemplified by the site for cleavage by EcoP15, the top strand of which reportedly reads CAGCAG. The bottom strand base complement, when read 5' to 3' is CTGCTG. Essentially by definition, restriction sites comprising odd numbers of bases (e.g., 5, 7) are non-palindromic. Certain endonucleases will cleave at variant forms of a site, which may be palindromic or not. For example, XbaI will recognize a site which reads (any purine)GATC(any pyrimidine) including the palindromic sequence AGATCT and the non-palindromic sequence GGATCT. Referring to the previously-noted "BRL Restriction Endonuclease Reference Chart," endonucleases recognizing six base palindromic sites exclusively include BbaI, Chul, Hin173, Ein91R, HinbIII, HinbIII, HindIII, Hinfl, Hsul, BglII, Stul, Rrul, Clal, AvaiII, Pvull, Smal, XmaI, Eccl, SacI, SboI, SbRI, ShyI, SstI, TglI, Avrl, Pvul, RshI, RspI, XbaI, XbaII, BglI, Msal, Scul, SexI, SgoI, SlaI, SliI, Spal, XbaI, XpaI, Bce170, Bsu1247, PstI, SalPI, XmaII, XorI, EcoRI, Rsh630I, SacI, SstI, SphI, BamHI, BamKI, BamNI, BamFI, BstI, KpnI, SalI, XmaI, HpaI, XbaI, AtuCI, BclI,

Cpel, SstIV, AosI, MstI, Ball, AsuII, and MlaI. Endonucleases which recognize only non-palindromic six base sequences exclusively include TthIII, EcoP15, Aval, and Avrl. Endonucleases recognizing both palindromic and non-palindromic six base sequences include HaeI, HgiAI, Acyl, AosII, AsuII, Accl, Chull, HinclI, HindII, MnnI, XholI, HaeII, HinHI, NgolI, and EcoRI'.

Upon determination of the structure of a desired polypeptide to be produced, practice of the present invention involves: preparation of two or more different specific, continuous double stranded DNA subunit sequences of 100 or more base pairs in length and having terminal portions of the proper configuration; serial insertion of subunits into a selected assembly vector with intermediate amplification of the hybrid vectors in a selected host organism; use of the assembly vector (or an alternate, selected "expression" vector including the DNA sequence which has been manufactured from the subunits) to transform a suitable, selected host; and, isolating polypeptide sequences expressed in the host organism. In its most efficient forms, practice of the invention involves using the same vector for assembly of the manufactured sequence and for large scale expression of the polypeptide. Similarly, the host microorganism employed for expression will ordinarily be the same as employed for amplifications performed during the subunit assembly process.

The manufactured DNA sequence may be provided with a promoter/regulator region for autonomous control of expression or may be incorporated into a vector in a manner providing for control of expression by a promoter/regulator sequence extant in the vector. Manufactured DNA sequences of the invention may suitably be incorporated into existing plasmid-borne genes (e.g., β -galactosidase) to form fusion genes coding for fusion polypeptide products including the desired amino acid sequences coded for by the manufactured DNA sequences.

In practice of the invention in its preferred forms, polypeptides produced may vary in size from about 65 or 70 amino acids up to about 200 or more amino acids. High levels of expression of the desired polypeptide by selected transformed host organisms is facilitated through the manufacture of DNA sequences which include one or more alternative codons which are preferentially expressed by the host.

Manufacture of double stranded subunit DNA sequences of 100 to 200 base pairs in length may proceed according to prior art assembly methods previously referred to, but is preferably accomplished by means of the rapid and efficient procedures disclosed in the aforementioned U.S. Application S.N. 375,493 by Stabinsky and used in certain of the following examples of actual practice of the present invention. Briefly put, these procedures involve the assembly from deoxyoligonucleotides of two or more different, linear, duplex DNA strands each including a relatively long double stranded region along with a relatively short single stranded region on one or both opposing ends of the double strand. The double stranded regions are designed to include codons needed to specify assembly of an initial, or terminal or intermediate portion of the total amino acid sequence of the desired polypeptide. Where possible, alternative codons preferentially expressed by a projected host (e.g., *E. coli*) are employed. Depending on the relative position to be assumed in the finally assembled subunit DNA sequence, the single stranded region(s) of the duplex strands will include a sequence of bases which, when complemented by bases of other duplex strands, also provide codons specifying amino acids within the desired polypeptide sequence.

Duplex strands formed according to this procedure are then enzymatically annealed to the one or two different duplex strands having complementary short, single stranded regions to form a desired continuous double stranded subunit DNA sequence which codes for the desired polypeptide fragment.

High efficiencies and rapidity in total sequence assembly are augmented in such procedures by performing a single annealing reaction involving three or more duplex strands, the short, single stranded regions of which constitute the base complement of at most one other single stranded region of any other duplex strand. Providing all duplex strands formed with short single stranded regions which uniquely complement only one of the single stranded regions of any other duplex is accomplished by alternative codon selection within the context of genetic code redundancy, and preferably also in the context of codon preferences of the projected host organism.

The following description of the manufacture of a hypothetical long DNA sequence coding for a hypothetical polypeptide will serve to graphically illustrate practice of the invention, especially in the context of formation of proper terminal sequences on subunit DNA sequences.

A biologically active polypeptide of interest is isolated and its amino acids are sequenced to reveal a constitution of 100 amino acid residues in a given continuous sequence. Formation of a manufactured gene for microbial expression of the polypeptide will thus require assembly of at least 300 base pairs for insertion into a selected viral or circular plasmid DNA vector to be used for transformation of a selected host organism.

A preliminary consideration in construction of the manufactured gene is the identity of the projected microbial host, because foreknowledge of the host allows for codon selection in the context of codon

preferences of the host species. For purposes of this discussion, the selection of an E. coli bacterial host is posited.

A second consideration in construction of the manufactured gene is the identity of the projected DNA vector employed in the assembly process. Selection of a suitable vector is based on existing knowledge of sites for cleavage of the vector by restriction endonuclease enzymes. More particularly, the assembly vector is selected on the basis of including DNA sequences providing endonuclease cleavage sites which will permit easy insertion of the subunits. In this regard, the assembly vector selected preferably has at least two restriction sites which occur only once (i.e., are "unique") in the vector prior to performance of any subunit insertion processes. For the purposes of this description, the selection of a hypothetical circular DNA plasmid pBR 3000 having a single EcoRI restriction site, i.e.,

-GAATTC-,
-CTTAA-

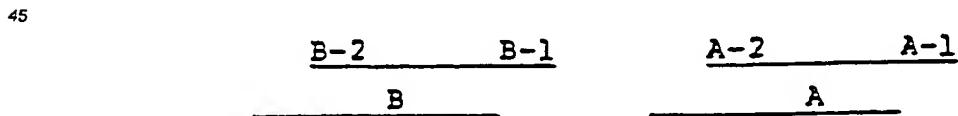
15 and a single PvuII restriction site, i.e.,

-CAGCTG-,
-GTCGAC-

is posited.

The amino acid sequence of the desired polypeptide is then analyzed in the context of determining availability of alternate codons for given amino acids (preferably in the context of codon preferences of the 25 projected E. coli host). With this information in hand, two subunit DNA sequences are designed, preferably having a length on the order of about 150 base pairs -- each coding for approximately one-half of the total amino acid sequences of the desired polypeptide. For purposes of this description, the two subunits manufactured will be referred to as "A" and "B".

The methods of the present invention as applied to two such subunits, generally call for: insertion of 30 one of the subunits into the assembly vector; amplification of the hybrid vector formed; and insertion of the second subunit to form a second hybrid including the assembled subunits in the proper sequence. Because the method involves joining the two subunits together in a manner permitting the joined ends to provide a continuous preselected sequence of bases coding for a continuous preselected sequence of amino acids, there exist certain requirements concerning the identity and sequence of the bases which make up the 35 terminal regions of the manufactured subunits which will be joined to another subunit. Because the method calls for joining subunits to the assembly vector, there exist other requirements concerning the identity and sequence of the bases which make up those terminal regions of the manufactured subunits which will be joined to the assembly vector. Because the subunits are serially, rather than concurrently, inserted into the assembly vector (and because the methods are most beneficially practiced when the subunits can be 40 selectively excised from assembled form to allow for alterations in selected base sequences therein), still further requirements exist concerning the identity of the bases in terminal regions of subunits manufactured. For ease of understanding in the following discussion of terminal region characteristics, the opposing terminal regions of subunits A and B are respectively referred to as A-1 and A-2, and B-1 and B-2, viz:



50 Assume that an assembly strategy is developed wherein subunit A is to be inserted into pBR3000 first, with terminal region A-1 to be ligated to the vector at the EcoRI restriction site. In the simplest case, the terminal region is simply provided with an EcoRI "sticky end", i.e., a single strand of four bases (-AATT- or -TTAA-) which will complement a single stranded sequence formed upon EcoRI digestion of pBR3000. This 55 will allow ligation of terminal region A-1 to the vector upon treatment with ligase enzyme. Unless the single strand at the end of terminal region A-1 is preceded by an appropriate base pair

(e.g., 5'-G-
 3'-CTTAA-),

5 the entire recognition site will not be reconstituted upon ligation to the vector. Whether or not the EcoRI
recognition site is reconstituted upon ligation (i.e., whether or not there will be 0 or 1 EcoRI sites remaining
after insertion of subunit A into the vector) is at the option of the designer of the strategy. Alternatively, one
10 may construct the terminal region A-1 of subunit A to include a complete set of base pairs providing a
recognition site for some other endonuclease, hypothetically designated "XXX", and then add on portions of
the EcoRI recognition site as above to provide an EcoRI "linker". To be of practical use in excising subunit
15 A from an assembled sequence, the "XXX" site should not appear elsewhere in the hybrid plasmid formed
upon insertion. The requirement for construction of terminal region A-1 is, therefore, that it comprise a
portion (i.e., all or part) of a base sequence which provides a recognition site for cleavage by a restriction
endonuclease, which recognition site is entirely present either once or not at all in the assembly vector
upon insertion of the subunit.

Assume that terminal region B-2 of subunit B is also to be joined to the assembly vector (e.g., at the
single recognition site for Pvull cleavage present on pBR3000). The requirements for construction of
20 terminal region B-2 are the same as for construction of A-1, except that the second endonuclease enzyme
in reference to which the construction of B-2 is made must be different from that with respect to which the
construction of A-1 is made. If recognition sites are the same, one will not be able to separately excise
segments A and B from the fully assembled sequence.

The above assumptions require, then, that terminal region A-2 is to be ligated to terminal region B-1 in
the final pBR3000 hybrid. Either the terminal region A-2 or the terminal region B-1 is constructed to
comprise a portion of a (preferably palindromic six base) recognition site for restriction endonuclease
25 cleavage by hypothetical third endonuclease "YYY" which recognition site will be entirely present once and
only once in the expression vector upon insertion of all subunits thereinto, i.e., at an intermediate position in
the assemblage of subunits. There exist a number of strategies for obtaining this result. In one alternative
strategy, the entire recognition site of "YYY" is contained in terminal region A-2 and the region additionally
30 includes the one or more portions of other recognition sites for endonuclease cleavage needed to (1)
complete the insertion of subunit A into the assembly vector for amplification purposes, and (2) allow for
subsequent joining of subunit A to subunit B. In this case, terminal region B-1 would have at its end only
the bases necessary to link it to terminal region A-2. In another alternative, the entire "YYY" recognition site
is included in terminal region B-1 and B-1 further includes at its end a portion of a recognition site for
endonuclease cleavage which is useful for joining subunit A to subunit B.

35 As another alternative, terminal region B-1 may contain at its end a portion of the "YYY" recognition
site. Terminal region A-2 would then contain the entire "YYY" recognition site plus, at its end, a suitable
"linker" for joining A-2 to the assembly vector prior to amplification of subunit A (e.g., a Pvull "sticky end").
After amplification of the hybrid containing subunit A, the hybrid would be cleaved with "YYY" (leaving a
40 sticky-ended portion of the "YYY" recognition site exposed on the end of A-2) and subunit B could be
inserted with its B-1 terminal region joined with the end of terminal region A-2 to reconstitute the entire
"YYY" recognition site. The requirement for construction of the terminal regions of all segments (other than
A-1 and B-2) is that one or the other or both (i.e., "at least half") comprise a portion (i.e., include all or part)
45 of a recognition site for third restriction endonuclease cleavage, which recognition site is entirely present
once and only once (i.e., is "unique") in said assembly vector after insertion of all subunits thereinto. To
generate a member of the class of novel DNA sequences of the invention, the recognition site of the third
endonuclease should be a six base palindromic recognition site.

While a subunit "terminal region" as referred to above could be considered to extend from the subunit
end fully halfway along the subunit to its center, as a practical matter the constructions noted would
ordinarily be performed in the final 10 or 20 bases. Similarly, while the unique "intermediate" recognition
50 site in the two subunit assemblage may be up to three times closer to one end of the manufactured
sequence than it is to the other, it will ordinarily be located near the center of the sequence. If, in the above
description, a synthetic plan was generated calling for preparation of three subunits to be joined, the
manufactured gene would include two unique restriction enzyme cleavage sites in intermediate positions at
least one of which will have a palindromic six base recognition site in the class of new DNA sequences of
55 the invention.

The significant advantages of the above-described process are manifest. Because the manufactured
gene now includes one or more unique restriction endonuclease cleavage sites at intermediate positions
along its length, modifications in the codon sequence of the two subunits joined at the cleavage site may be

effected with great facility and without the need to re-synthesize the entire manufactured gene.

Following are illustrative examples of the actual practice of the invention in formation of manufactured genes capable of directing the synthesis of: human leukocyte interferon of the F subtype (INF- α F) and analogs thereof; and, multiple consensus leukocyte interferons which, due to homology to IFN- α F can be 5 named as IFN- α F analogs. It will be apparent from these examples that the gene manufacturing methodology of the present invention provides an overall synthetic strategy for the truly rapid, efficient synthesis and expression of genes of a length in excess of 200 base pairs within a highly flexible framework allowing for variations in the structures of products to be expressed which has not heretofore been available to investigators practicing recombinant DNA techniques.

10

EXAMPLE 1

The amino acid sequences for the human leukocyte interferon of the F subtype has been deduced by way of sequencing of cDNA clones. See, e.g., Goedell, et al., *Nature*, 200, pp. 20-26 (1981). The general 15 procedures of prior Examples 1, 2 and 3 were employed in the design and assembly of a manufactured DNA sequence for use in microbial expression of IFN- α F in *E. coli* by means of a pBR322-derived expression vector. A general plan for the construction of three "major" subunit DNA sequences (LeuIFN-F I, LeuIFN-F II and LeuIFN-F III) and one "minor" subunit DNA sequence (LeuIFN-F IV) was evolved and is shown in Table IV below.

20

25

30

35

40

45

50

55

5

30

75

20

25

30

35

40

45

50

55

TABLE IV

Leuven-P IV

xbaI	-1	1	9	BstE II						
	Met	Cys	Asp	Leu	Pro	Gln	Thr	His	Ser	Leu
	<pre> CTAGAGA ATG TGT GAT TTA CCT CAA ACT CAT TCT CTT G TCT TAC ACA CTA ATT GGA GTT TGA GTA AGA GAA CCATTG 43 41 42 44 </pre>									

Leu IFN-F III

ECORI 10

Gly Asn Arg Arg Ala Leu Ile Leu Leu Ala Gln Met Gly Arg Ile Ser Pro Phe

AAATTCATTA	CCT AAC CGT CGC GCT CTG ATT CTG CTC GCA CAG ATG GGT CCT	37
CTAAAT CCA TTTG	GCA GCG CGA GAC TAA GAC GTC TAC CCA GCA TAA AGG GGC AAA	38
		39

BstE II

50

Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe

AGC TGC CTG AAA GAC CGT CAC GAC	TTC GGC TTT CCG CAA GAA GAG TTC GAT GGC TTT TCC	33
TCG ACG GAC TTT CTG GCA GTG CTG	CTG AAG CCG AAA GGC GTT CTT CTC AGC CTA CCG TTG GTT AAG	34
		35

50

Gln Lys Ala Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr Phe Asn Leu Phe Ser

CAG AAA GCT CAG GCA ATC TCT GAA CTG CAC GAA ATG ATC CAA CAG ACC TTC AAC CTG TTT TCC	29	
GTC TTT CGA GTC CGT TAG AGA CAA GAC GTC CTT TAC TAG GTT GTC TGG AGG TTG GAC AAA AGG	30	
		31

70

Thr Lys Asp Ser Ser Ala Thr Trp Glu Gln

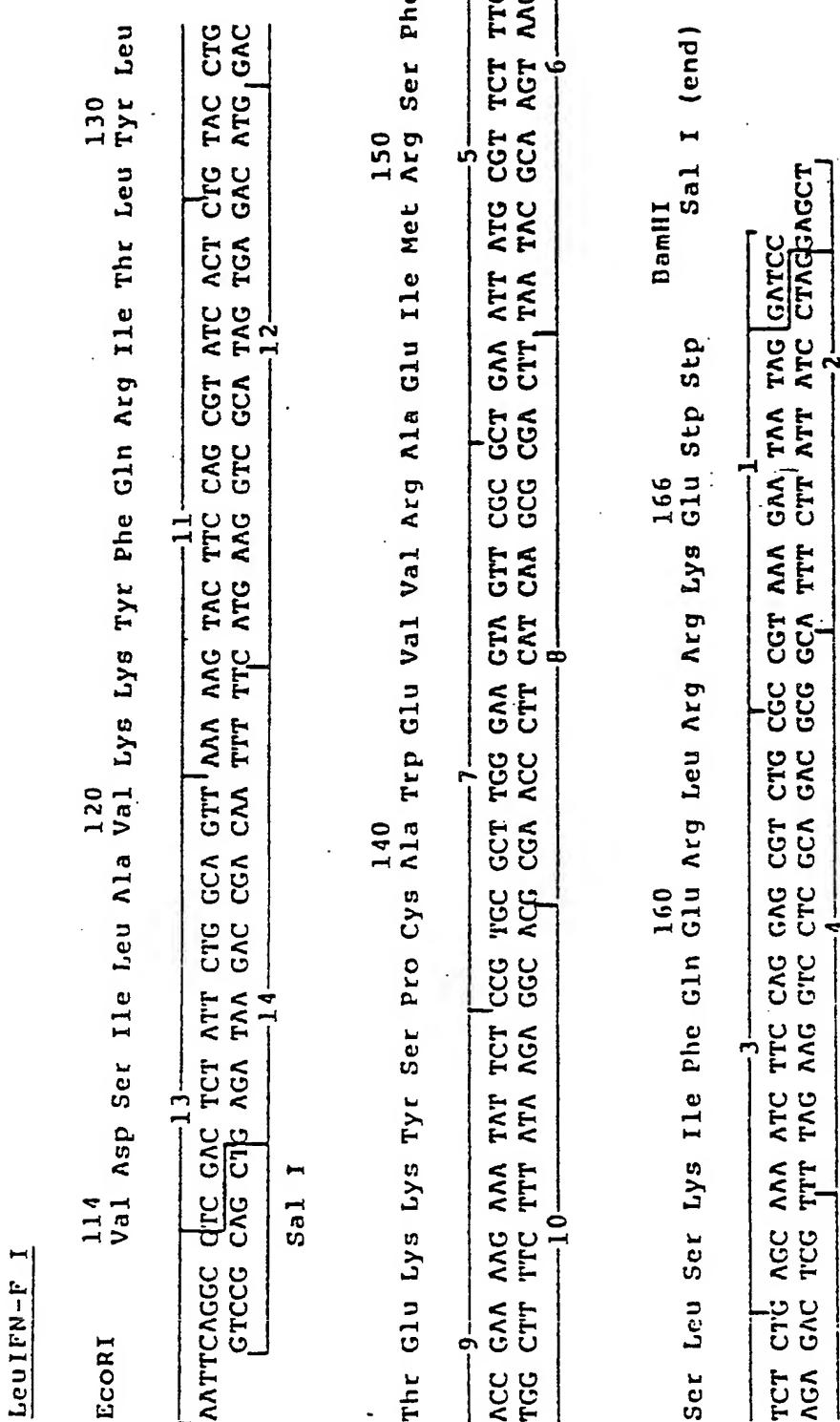
ACT AAA GAC AGC TCT GCT ACC TGG GAA CAA	25	
TGA TTT CTG TCG AGA CGA TGG ACC CTT GTT TCCAA	26	
		27

70

Ile Gln Gln Gln Thr Phe Asn Leu Phe Ser

5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95

LeuIFN- β II



"sticky end" but, upon complementation, a Sall recognition site is not reconstituted. A full BamHI recognition site remains, however, allowing for subsequent excision of the subunit.) This first intermediate plasmid is amplified and subunit II is inserted into the amplified plasmid after again cutting with EcoRI and Sall. The second intermediate plasmid thus formed is amplified and subunit III is inserted into the amplified 5 plasmid cut with EcoRI and HindIII. The third intermediate plasmid thus formed is amplified. Subunit IV is ligated to an EcoRI and XbaI fragment isolated from pINT_γ-TXb4 of Example 4 and this ligation product (having EcoRI and BstEII sticky ends) is then inserted into the amplified third intermediate plasmid cut with EcoRI and BstEII to yield the final expression vector.

The isolated product of trp promoter/operator controlled E.coli expression of the manufactured DNA 10 sequence of Table IV as inserted into the final expression vector was designated IFN- α F₁.

EXAMPLE 2

As discussed infra with respect to consensus leukocyte interferon, those human leukocyte interferon 15 subtypes having a threonine residue at position 14 and a methionine residue at position 16 are reputed to display greater antiviral activity than those subtypes possessing Ala¹⁴ and Ile¹⁶ residues. An analog of human leukocyte interferon subtype F was therefore manufactured by means of microbial expression of a DNA sequence of Example 1 which had been altered to specify threonine and methionine as residues 14 and 16, respectively. More specifically, [Thr¹⁴, Met¹⁶] IFN- α F, designated IFN- α F₂, was expressed in E.coli 20 upon transformation with a vector of Example 1 which had been cut with Sall and HindIII and into which a modified subunit II (of Table VII) was inserted. The specific modifications of subunit II involved assembly with segment 39 altered to replace the alanine-specifying codon, GCT, with a threonine-specifying ACT codon and replace the isoleucine-specifying codon, ATT, with an ATG codon. Corresponding changes in complementary bases were made in section 40 of subunit LeuIFN-FII.

The following Examples 2 and 3 relate to practice of the invention in the microbial synthesis of 25 consensus human leukocyte interferon polypeptides which can be designated as analogs of human leukocyte interferon subtype F.

EXAMPLE 3

"Consensus human leukocyte interferon" ("IFN-Con," "LeuIFN-Con") as employed herein shall mean a 30 non-naturally-occurring polypeptide which predominantly includes those amino acid residues which are common to all naturally-occurring human leukocyte interferon subtype sequences and which includes, at one or more of those positions wherein there is no amino acid common to all subtypes, an amino acid 35 which predominantly occurs at that position and in no event includes any amino acid residue which is not extant in that position in at least one naturally-occurring subtype. (For purposes of this definition, subtype A is positionally aligned with other subtypes and thus reveals a "missing" amino acid at position 44.) As so defined, a consensus human leukocyte interferon will ordinarily include all known common amino acid residues of all subtypes. It will be understood that the state of knowledge concerning naturally-occurring 40 subtype sequences is continuously developing. New subtypes may be discovered which may destroy the "commonality" of a particular residue at a particular position. Polypeptides whose structures are predicted on the basis of a later-amended determination of commonality at one or more positions would remain within the definition because they would nonetheless predominantly include common amino acids and because those amino acids no longer held to be common would nonetheless quite likely represent the predominant 45 amino acid at the given positions. Failure of a polypeptide to include either a common or predominant amino acid at any given position would not remove the molecule from the definition so long as the residue at the position occurred in at least one subtype. Polypeptides lacking one or more internal or terminal residues of consensus human leukocyte interferon or including internal or terminal residues having no counterpart in any subtype would be considered analogs of human consensus leukocyte interferon.

Published predicted amino acid sequences for eight cDNA-derived human leukocyte interferon subtypes were analyzed in the context of the identities of amino acids within the sequence of 166 residues. See, generally, Goedell, et al., Nature, 290, pp. 20-26 (1981) comparing LeIFN-A through LeIFN-H and noting that only 79 amino acids appear in identical positions in all eight interferon forms and 99 amino acids appear in identical positions if the E subtype (deduced from a cDNA pseudogene) was ignored. Each of the 55 remaining positions was analyzed for the relative frequency of occurrence of a given amino acid and, where a given amino acid appeared at the same position in at least five of the eight forms, it was designated as the predominant amino acid for that position. A "consensus" polypeptide sequence of 166 amino acids was plotted out and compared back to the eight individual sequences, resulting in the determination that LeIFN-

F required few modifications from its "naturally-occurring" form to comply with the consensus sequence.

A program for construction of a manufactured IFN-Con DNA sequence was developed and is set out below in Table V. In the table, an asterisk designates the variations in IFN- α F needed to develop LeIFN-Con₁, i.e., to develop the [Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸] analog of IFN- α F. The illustrated top strand sequence includes, wherever possible, codons noted to be the subject of preferential expression in E. coli. The sequence also includes bases providing recognition sites for Sal, HindIII, and BstE2 at positions intermediate the sequence and for XbaI and BamHI at its ends. The latter sites are selected for use in incorporation of the sequence in a pBR322 vector, as was the case with the sequence developed for IFN- α F and its analogs.

10

15

20

25

30

35

40

45

50

55

TABLE V

5 -1 1 10
 Met-Cys-Asp-Leu-Pro-Gln-Thr-His-Ser-Leu-Gly-Asn-Arg-Arg-
 ATG TGT GAT TTA CCT CAA ACT CAT TCT CTT GGT AAC CGT CGC
 10 20 *
 Ala-Leu-Ile-Leu-Leu-Ala-Gln-Met-Arg-Arg-Ile-Ser-Pro-Phe-
 GCT CTG ATT CTG CTG GCA CAG ATG CGT CGT ATT TCC CCG TTT
 15 30 40
 Ser-Cys-Leu-Lys-Asp-Arg-His-Asp-Phe-Gly-Phe-Pro-Gln-Glu-
 AGC TGC CTG AAA GAC CGT CAC GAC TTC GGC TTT CCG CAA GAA
 20 50
 Glu-Phe-Asp-Gly-Asn-Gln-Phe-Gln-Lys-Ala-Gln-Ala-Ile-Ser-
 GAG TTC GAT GGC AAC CAA TTC CAG AAA GCT CAG GCA ATC TCT
 25 60
 Val-Leu-His-Glu-Met-Ile-Gln-Gln-Thr-Phe-Asn-Leu-Phe-Ser-
 GTA CTG CAC GAA ATG ATC CAA CAG ACC TTC AAC CTG TTT TCC
 30 70 * * * 80
 Thr-Lys-Asp-Ser-Ser-Ala-Ala-Trp-Asp-Glu-Ser-Leu-Leu-Glu-
 ACT AAA GAC AGC TCT GCT TGG GAC GAA AGC TTG CTG GAG
 35 * * 90 *
 Lys-Phe-Tyr-Thr-Glu-Leu-Tyr-Gln-Gln-Leu-Asn-Asp-Leu-Glu-
 AAG TTC TAC ACT GAA CTG TAT CAG CAG CTG AAC GAC CTG GAA
 40 100 110
 Ala-Cys-Val-Ile-Gln-Glu-Val-Gly-Val-Glu-Glu-Thr-Pro-Leu-
 GCA TGC GTA ATC CAG GAA GTT GGT GTA GAA GAG ACT CCG CTG
 45 120
 Met-Asn-Val-Asp-Ser-Ile-Leu-Ala-Val-Lys-Lys-Tyr-Phe-Gln-
 ATG AAC GTC GAC TCT ATT CTG GCA GTT AAA AAG TAC TTC CAG
 50 130
 Arg-Ile-Thr-Leu-Tyr-Leu-Thr-Glu-Lys-Lys-Tyr-Ser-Pro-Cys-
 CGT ATC ACT CTG TAC CTG ACC GAA AAG AAA TAT TCT CCG TGC
 55 140 150
 Ala-Trp-Glu-Val-Val-Arg-Ala-Glu-Ile-Met-Arg-Ser-Phe-Ser-
 GCT TGG GAA GTA GTT CGC GCT GAA ATT ATG CGT TCT TTC TCT
 60 * * * 160 166 Stop
 Leu-Ser-Thr-Asn-Leu-Gln-Glu-Arg-Leu-Arg-Arg-Lys-Glu-
 CTG TCT ACT AAC CTG CAG GAG CGT CTG CGC CGT AAA GAA TAA
 65 Stop
 70 TAG

Table VI below sets out the specific double stranded DNA sequence for preparation 4 subunit DNA sequences for use in manufacture of IFN-Con₁. Subunit LeuIFN-Con IV is a duplicate of LeuIFN-F IV of Table V. Segments of subunits which differ from those employed to construct the IFN- α F gene are designated with a "prime" (e.g., 37' and 38' are altered forms of sections 37 and 38 needed to provide arginine rather than glycine at position 22).

5

70

75

20

25

30

35

40

45

50

55

TABLE VI

LeuIFN Con IV

XbaI	-1	1	2	3	4	5	6	7	8	9	BstE II
	Met	Cys	Asp	Leu	Pro	Gln	Thr	His	Ser	Leu	
	CTA GAG GAT TGT GAT TTA CCT CAA ACT CAT TCT CTT G TCT TAC ACA CTA ATT GGA GTT TGA GAA AGA GAA CCATTG 43 44 45 46 47 48 49 50 51 52 53 42										

50 45 40 35 30 25 20 15 10 5

LeuIFN Con III

ECORI 10

Gly Asn Arg Arg Ala Leu Ile Leu Ala Leu Ile Met Arg Arg Ile Ser Pro Phe

39 37' 39
 ATTCACTTA CCT AAC CGC CGT C'GT ATT CTG CTG GCA CAG ATG CGT ATT TCC CCG TTT
 GTAAAT CCA 'TTG GCA CGG GAC TAA GAC GAC CGT GTC TAC GCA GCA TAA AGG GGC AAA
 40 38',

BstE II

30 40 30 35 33' 33
 Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe

35 35 35 33' 33
 AGC TGC CTG AAA GAC CGT CAC GAC T'GC GGC TTG CCG CAA GAA GAG TTC GAT GGC AAC CAA TTC
 TCG ACG GAC TTT CTG GCA GTG CTG AAA GGC GTT CTT CTC AGC CTA CCG TTT GTT AAG
 36 34',

50 60 50 55 29 29
 Gln Lys Ala Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr Phe Asn Leu Phe Ser

31 31 31 30 29 29
 CAG AAA CCT' CAG GCA ATC TCT GTC CTC CAC GAA ATG ATC CAA CAG ACC ATC AAC CTC TTT TCC
 GTC 'TTT CGA GTC CGT TAG AGA CTT GAC GTC CTT TAC TAG GTG GTC TGG AAA TGT GAC AAA AGG
 32 30',

70 79 70 75 27 27
 Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu

25' 25' 25' 25' 25' 25'
 AC' 'AAA GAC AGC AGC TCT GCT GCT GCG TGG GAC GAA
 TGA TTT CTG TCG AGA CGA CGA ACC CTC CTT TCG
 26',

Lcu IFN Con II

Leu Asn Asp Leu Glu Ala Cys Val Ile Gln Glu Val Gly Val Glu Glu
 100
 19' _____ 17' _____ 18' _____
 CTC AAC GAC CTG GAA GCA TGC GCA ATC CAG GAA CTT GGT GTA GAA GAG
 GAC T'RG C'RG GAC C'TT' CGT' ACG' CAT' TAC' GTC' CTT' CAA' CCA' CAT' CTT' CTC'

5
10
15
20
25
30
35
40
45
50LeuIFN Con 1

E_{COLI} 114 Val Asp Ser Ile Leu Ala Val Lys Tyr Phe Gln Arg Ile Thr Leu Tyr Leu
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995
 1000
 1005
 1010
 1015
 1020
 1025
 1030
 1035
 1040
 1045
 1050
 1055
 1060
 1065
 1070
 1075
 1080
 1085
 1090
 1095
 1100
 1105
 1110
 1115
 1120
 1125
 1130
 1135
 1140
 1145
 1150
 1155
 1160
 1165
 1170
 1175
 1180
 1185
 1190
 1195
 1200
 1205
 1210
 1215
 1220
 1225
 1230
 1235
 1240
 1245
 1250
 1255
 1260
 1265
 1270
 1275
 1280
 1285
 1290
 1295
 1300
 1305
 1310
 1315
 1320
 1325
 1330
 1335
 1340
 1345
 1350
 1355
 1360
 1365
 1370
 1375
 1380
 1385
 1390
 1395
 1400
 1405
 1410
 1415
 1420
 1425
 1430
 1435
 1440
 1445
 1450
 1455
 1460
 1465
 1470
 1475
 1480
 1485
 1490
 1495
 1500
 1505
 1510
 1515
 1520
 1525
 1530
 1535
 1540
 1545
 1550
 1555
 1560
 1565
 1570
 1575
 1580
 1585
 1590
 1595
 1600
 1605
 1610
 1615
 1620
 1625
 1630
 1635
 1640
 1645
 1650
 1655
 1660
 1665
 1670
 1675
 1680
 1685
 1690
 1695
 1700
 1705
 1710
 1715
 1720
 1725
 1730
 1735
 1740
 1745
 1750
 1755
 1760
 1765
 1770
 1775
 1780
 1785
 1790
 1795
 1800
 1805
 1810
 1815
 1820
 1825
 1830
 1835
 1840
 1845
 1850
 1855
 1860
 1865
 1870
 1875
 1880
 1885
 1890
 1895
 1900
 1905
 1910
 1915
 1920
 1925
 1930
 1935
 1940
 1945
 1950
 1955
 1960
 1965
 1970
 1975
 1980
 1985
 1990
 1995
 2000
 2005
 2010
 2015
 2020
 2025
 2030
 2035
 2040
 2045
 2050
 2055
 2060
 2065
 2070
 2075
 2080
 2085
 2090
 2095
 2100
 2105
 2110
 2115
 2120
 2125
 2130
 2135
 2140
 2145
 2150
 2155
 2160
 2165
 2170
 2175
 2180
 2185
 2190
 2195
 2200
 2205
 2210
 2215
 2220
 2225
 2230
 2235
 2240
 2245
 2250
 2255
 2260
 2265
 2270
 2275
 2280
 2285
 2290
 2295
 2300
 2305
 2310
 2315
 2320
 2325
 2330
 2335
 2340
 2345
 2350
 2355
 2360
 2365
 2370
 2375
 2380
 2385
 2390
 2395
 2400
 2405
 2410
 2415
 2420
 2425
 2430
 2435
 2440
 2445
 2450
 2455
 2460
 2465
 2470
 2475
 2480
 2485
 2490
 2495
 2500
 2505
 2510
 2515
 2520
 2525
 2530
 2535
 2540
 2545
 2550
 2555
 2560
 2565
 2570
 2575
 2580
 2585
 2590
 2595
 2600
 2605
 2610
 2615
 2620
 2625
 2630
 2635
 2640
 2645
 2650
 2655
 2660
 2665
 2670
 2675
 2680
 2685
 2690
 2695
 2700
 2705
 2710
 2715
 2720
 2725
 2730
 2735
 2740
 2745
 2750
 2755
 2760
 2765
 2770
 2775
 2780
 2785
 2790
 2795
 2800
 2805
 2810
 2815
 2820
 2825
 2830
 2835
 2840
 2845
 2850
 2855
 2860
 2865
 2870
 2875
 2880
 2885
 2890
 2895
 2900
 2905
 2910
 2915
 2920
 2925
 2930
 2935
 2940
 2945
 2950
 2955
 2960
 2965
 2970
 2975
 2980
 2985
 2990
 2995
 3000
 3005
 3010
 3015
 3020
 3025
 3030
 3035
 3040
 3045
 3050
 3055
 3060
 3065
 3070
 3075
 3080
 3085
 3090
 3095
 3100
 3105
 3110
 3115
 3120
 3125
 3130
 3135
 3140
 3145
 3150
 3155
 3160
 3165
 3170
 3175
 3180
 3185
 3190
 3195
 3200
 3205
 3210
 3215
 3220
 3225
 3230
 3235
 3240
 3245
 3250
 3255
 3260
 3265
 3270
 3275
 3280
 3285
 3290
 3295
 3300
 3305
 3310
 3315
 3320
 3325
 3330
 3335
 3340
 3345
 3350
 3355
 3360
 3365
 3370
 3375
 3380
 3385
 3390
 3395
 3400
 3405
 3410
 3415
 3420
 3425
 3430
 3435
 3440
 3445
 3450
 3455
 3460
 3465
 3470
 3475
 3480
 3485
 3490
 3495
 3500
 3505
 3510
 3515
 3520
 3525
 3530
 3535
 3540
 3545
 3550
 3555
 3560
 3565
 3570
 3575
 3580
 3585
 3590
 3595
 3600
 3605
 3610
 3615
 3620
 3625
 3630
 3635
 3640
 3645
 3650
 3655
 3660
 3665
 3670
 3675
 3680
 3685
 3690
 3695
 3700
 3705
 3710
 3715
 3720
 3725
 3730
 3735
 3740
 3745
 3750
 3755
 3760
 3765
 3770
 3775
 3780
 3785
 3790
 3795
 3800
 3805
 3810
 3815
 3820
 3825
 3830
 3835
 3840
 3845
 3850
 3855
 3860
 3865
 3870
 3875
 3880
 3885
 3890
 3895
 3900
 3905
 3910
 3915
 3920
 3925
 3930
 3935
 3940
 3945
 3950
 3955
 3960
 3965
 3970
 3975
 3980
 3985
 3990
 3995
 4000
 4005
 4010
 4015
 4020
 4025
 4030
 4035
 4040
 4045
 4050
 4055
 4060
 4065
 4070
 4075
 4080
 4085
 4090
 4095
 4100
 4105
 4110
 4115
 4120
 4125
 4130
 4135
 4140
 4145
 4150
 4155
 4160
 4165
 4170
 4175
 4180
 4185
 4190
 4195
 4200
 4205
 4210
 4215
 4220
 4225
 4230
 4235
 4240
 4245
 4250
 4255
 4260
 4265
 4270
 4275
 4280
 4285
 4290
 4295
 4300
 4305
 4310
 4315
 4320
 4325
 4330
 4335
 4340
 4345
 4350
 4355
 4360
 4365
 4370
 4375
 4380
 4385
 4390
 4395
 4400
 4405
 4410
 4415
 4420
 4425
 4430
 4435
 4440
 4445
 4450
 4455
 4460
 4465
 4470
 4475
 4480
 4485
 4490
 4495
 4500
 4505
 4510
 4515
 4520
 4525
 4530
 4535
 4540
 4545
 4550
 4555
 4560
 4565
 4570
 4575
 4580
 4585
 4590
 4595
 4600
 4605
 4610
 4615
 4620
 4625
 4630
 4635
 4640
 4645
 4650
 4655
 4660
 4665
 4670
 4675
 4680
 4685
 4690
 4695
 4700
 4705
 4710
 4715
 4720
 4725
 4730
 4735
 4740
 4745
 4750
 4755
 4760
 4765
 4770
 4775
 4780
 4785
 4790
 4795
 4800
 4805
 4810
 4815
 4820
 4825
 4830
 4835
 4840
 4845
 4850
 4855
 4860
 4865
 4870
 4875
 4880
 4885
 4890
 4895
 4900
 4905
 4910
 4915
 4920
 4925
 4930
 4935
 4940
 4945
 4950
 4955
 4960
 4965
 4970
 4975
 4980
 4985
 4990
 4995
 5000
 5005
 5010
 5015
 5020
 5025
 5030
 5035
 5040
 5045
 5050
 5055
 5060
 5065
 5070
 5075
 5080
 5085
 5090
 5095
 5100
 5105
 5110
 5115
 5120
 5125
 5130
 5135
 5140
 5145
 5150
 5155
 5160
 5165
 5170
 5175
 5180
 5185
 5190
 5195
 5200
 5205
 5210
 5215
 5220
 5225
 5230
 5235
 5240
 5245
 5250
 5255
 5260
 5265
 5270
 5275
 5280
 5285
 5290
 5295
 5300
 5305
 5310
 5315
 5320
 5325
 5330
 5335
 5340
 5345
 5350
 5355
 5360
 5365
 5370
 5375
 5380
 5385
 5390
 5395
 5400
 5405
 5410
 5415
 5420
 5425
 5430
 5435
 5440
 5445
 5450
 5455
 5460
 5465
 5470
 5475
 5480
 5485
 5490
 5495
 5500
 5505
 5510
 5515
 5520
 5525
 5530
 5535
 5540
 5545
 5550
 5555
 5560
 5565
 5570
 5575
 5580
 5585
 5590
 5595
 5600
 5605
 5610
 5615
 5620
 5625
 5630
 5635
 5640
 5645
 5650
 5655
 5660
 5665
 5670
 5675
 5680
 5685
 5690
 5695
 5700
 5705
 5710
 5715
 5720
 5725
 5730
 5735
 5740
 5745
 5750
 5755
 5760
 5765
 5770
 5775
 5780
 5785
 5790
 5795
 5800
 5805
 5810
 5815
 5820
 5825
 5830
 5835
 5840
 5845
 5850
 5855
 5860
 5865
 5870
 5875
 5880
 5885
 5890
 5895
 5900
 5905
 5910
 5915
 5920
 5925
 5930
 5935
 5940
 5945
 5950
 5955
 5960
 5965
 5970
 5975
 5980
 5985
 5990
 5995
 6000
 6005
 6010
 6015
 6020
 6025
 6030
 6035
 6040
 6045
 6050
 6055
 6060
 6065
 6070
 6075
 6080
 6085
 6090
 6095
 6100
 6105
 6110
 6115
 6120
 6125
 6130
 6135
 6140
 6145
 6150
 6155
 6160
 6165
 6170
 6175
 6180
 6185
 6190
 6195
 6200
 6205
 6210
 6215
 6220
 6225
 6230
 6235
 6240
 6245
 6250
 6255
 6260
 6265
 6270
 6275
 6280
 6285
 6290
 6295
 6300
 6305
 6310
 6315
 6320
 6325
 6330
 6335
 6340
 6345
 6350
 6355
 6360
 6365
 6370
 6375
 6380
 6385
 6390
 6395
 6400
 6405
 6410
 6415
 6420
 6425
 6430
 6435
 6440
 6445
 6450
 6455
 6460
 6465
 6470
 6475
 6480
 6485
 6490
 6495
 6500
 6505
 6510
 6515
 6520
 6525
 6530
 6535
 6540
 6545
 6550
 6555
 6560
 6565
 6570
 6575
 6580
 6585
 6590
 6595
 6600
 6605
 6610
 6615
 6620
 6625
 6630
 6635
 6640
 6645
 6650
 6655
 6660
 6665
 6670
 6675
 6680
 6685
 6690
 6695
 6700
 6705
 6710
 6715
 6720
 6725
 6730
 6735
 6740
 6745
 6750
 6755
 6760
 6765
 6770
 6775
 6780
 6785
 6790
 6795
 6800
 6805
 6810
 6815
 6820
 6825
 6830
 6835
 6840
 6845
 6850
 6855
 6860
 6865
 6870
 6875
 6880
 6885
 6890
 6895
 6900
 6905
 6910
 6915
 6920
 6925
 6930
 6935
 6940
 6945
 6950
 6955
 6960
 6965
 6970
 6975
 6980
 6985
 6990
 6995
 7000
 7005
 7010
 7015
 7020
 7025
 7030
 7035
 7040
 7045
 7050
 7055
 7060
 7065
 7070
 7075
 7080
 7085
 7090
 7095
 7100
 7105
 7110
 7115
 7120
 7125
 7130
 7135
 7140
 7145
 7150
 7155
 7160
 7165
 7170
 7175
 7180
 7185
 7190
 7195
 7200
 7205
 7210<br

the reference's summary of sequences. The four above-noted residues were retained from the native IFN- α F sequence to facilitate construction of subunits and assembly of subunits into an expression vector. (Note, e.g., serine was retained at position 80 to allow for construction of a HindIII site.)

Since publication of the Goedell, et al. summary of IFN- α subtypes, a number of additional subtypes have been ascertained. Figure 1 sets out in tabular form the deduced sequences of the 13 presently known subtypes (exclusive of those revealed by five known cDNA pseudogenes) with designations of the same IFN- α subtypes from different laboratories indicated parenthetically (e.g., IFN- α 6 and IFN- α K). See, e.g., Goedell, et al., *supra*; Stebbing, et al., in: *Recombinant DNA Products, Insulin, Interferons and Growth Hormones* (A. Bollen, ed.), CRC Press (1983); and Weissman, et al., *U.C.L.A. Symp.Mol.Cell Biol.*, 25, pp. 295-326 (1982). Positions where there is no common amino acid are shown in bold face. IFN- α subtypes are roughly grouped on the basis of amino acid residues. In seven positions (14, 16, 71, 78, 79, 83, and 160) the various subtypes show just two alternative amino acids, allowing classification of the subtypes into two subgroups (I and II), based on which of the seven positions are occupied by the same amino acid residues. Three IFN- α subtypes (H, F and B) cannot be classified as Group I or Group II and, in terms of distinguishing positions, they appear to be natural hybrids of both group subtypes. It has been reported that IFN- α subtypes of the Group I type display relatively high antiviral activity while those of Group II display relatively high antitumor activity.

IFN-Con₁ structure is described in the final line of the Figure. It is noteworthy that certain residues of IFN-Con₁ (e.g., serine at position 8) which were determined to be "common" on the basis of the Goedell, et al., sequences are now seen to be "predominant". Further, certain of the IFN-Con₁ residues determined to be predominant on the basis of the reference (Arg²², Asp⁷⁸, Glu⁷⁹, and Tyr⁸⁶) are no longer so on the basis of updated information, while certain heretofore non-predominant others (Ser⁸⁰ and Glu⁸³) now can be determined to be predominant.

25 EXAMPLE 4

A human consensus leukocyte interferon which differed from IFN-Con₁ in terms of the identity of amino acid residues at positions 14 and 16 was prepared by modification of the DNA sequence coding for IFN-Con₁. More specifically, the expression vector for IFN-Con₁ was treated with BstEII and Hind III to delete subunit LeuIFN Con III. A modified subunit was inserted wherein the alanine-specifying codon, GCT, of sections 39 and 40 was altered to a threonine-specifying codon, ACT, and the isoleucine codon, CTG, was changed to ATG. The product of expression of the modified manufactured gene, [Thr¹⁴, Met¹⁶, Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁸⁰, Leu⁸⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸]IFN- α F, was designated IFN-Con₂.

Presently being constructed is a gene for a consensus human leukocyte interferon polypeptide which will differ from IFN-Con₁ in terms of the identity of residues at positions 114 and 121. More specifically, the Val¹¹⁴ and Lys¹²¹ residues which duplicate IFN- α F subtype residues but are not predominant amino acids will be changed to the predominant Glu¹¹⁴ and Arg¹²¹ residues, respectively. Because the codon change from Val¹¹⁴ to Arg¹¹⁴ (e.g., GTC to GAA) will no longer allow for a Sall site at the terminal portion of subunit LeuIFN Con I (of Table VI), subunits I and II will likely need to be constructed as a single subunit. Changing the AAA, lysine, codon of sections 11 and 12 to CTG will allow for the presence of arginine at position 121. The product of microbial expression of the manufactured gene, [Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁸⁰, Leu⁸⁶, Glu¹¹⁴, Arg¹²¹, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸] IFN- α F, will be designated IFN-Con₃.

The following example relates to antiviral activity screening of human leukocyte interferon and polypeptides provided by the preceding examples.

45 EXAMPLE 5

Table VII below provides the results of testing of antiviral activity in various cell lines of natural (buffy coat) interferon and isolated, microbially-expressed, polypeptides designated IFN- α F₁, IFN- α F₂, IFN-Con₁, and IFN-Con₂. Viruses used were VSV (vesicular stomatitis virus) and EMCV (encephalomyocarditis virus). Cell lines were from various mammalian sources, including human (WISH, HeLa), bovine (MDBK), mouse (MLV-6), and monkey (Vero). Antiviral activity was determined by an end-point cytopathic effect assay as described in Weck, et al., *J.Gen.Viro.*, 57, pp. 233-237 (1981) and Campbell, et al., *Can.J.Microbiol.*, 21, pp. 1247-1253 (1975). Data shown was normalized for antiviral activity in WISH cells.

TABLE VII

5	<u>Virus</u>	<u>Cell Line</u>	<u>Buffy Coat</u>	<u>IFN- αF₁</u>	<u>IFN- αF₂</u>	<u>IFN- Con₁</u>	<u>IFN- Con₂</u>
10	VSV	WISH	100	100	100	100	100
VSV	HeLa	400	100	ND*	200	100	
15	VSV	MDBK	1600	33	ND	200	300
VSV	MLV-6	20	5	ND	3	20	
20	VSV	Vero	10	0.1	ND	10	0.1
EMCV	WISH	100	100	100	100	100	
EMCV	HeLa	100	5	ND	33	33	
EMCV	Vero	100	20	ND	1000	10	

*ND - no data presently available.

It will be apparent from the above examples that the present invention provides, for the first time, an entire new genus of synthesized, biologically active proteinaceous products which products differ from naturally-occurring forms in terms of the identity and/or location of one or more amino acids and in terms of one or more biological (e.g., antibody reactivity) and pharmacological (e.g., potency or duration of effect) but which substantially retain other such properties.

Products of the present invention and/or antibodies thereto may be suitably "tagged", for example radiolabelled (e.g., with I¹²⁵) conjugated with enzymes or fluorescently labelled, to provide reagent materials useful in assays and/or diagnostic test kits, for the qualitative and/or quantitative determination of the presence of such products and/or said antibodies in fluid samples. Such antibodies may be obtained from the innoculation of one or more animal species (e.g., mice rabbit, goat, human, etc.) or from monoclonal antibody sources. Any of such reagent materials may be used alone or in combination with a suitable substrate, e.g., coated on a glass or plastic particle bead.

Numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the foregoing illustrative examples. Consequently, the invention should be considered as limited only to the extent reflected by the appended claims.

40 Claims

Claims for the following Contracting States : BE, CH, DE, FR, GB, LI, LU, NL, SE

1. A consensus leukocyte interferon protein predominantly including those amino acids which are common to all naturally-occurring human alpha IFN subtype sequences and including at one or more positions where there is no amino acid common to all subtypes an amino acid which predominantly occurs at that position and in no event including any amino acid residue which is not extant at that position in at least one naturally-occurring subtype.
2. A consensus human leukocyte interferon protein according to Claim 1 selected from the group consisting of:

[Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
 Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu<sup>158}] IFN- α F;
 [Thr¹⁴, Met¹⁶, Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹,
 Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu<sup>158}] IFN- α F; and
 [Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
 Leu⁹⁶, Glu¹¹⁴, Arg¹²¹, Thr¹⁵⁶, Asn¹⁵⁷, Leu^{158}] IFN- α F.}</sup></sup>

3. A gene capable of directing the synthesis in a selected host microorganism of consensus human leukocyte interferon protein according to Claim 1 or 2.

Claims for the following Contracting State : AT

25 1. Product produced by a microbiological process, comprising a consensus leukocyte interferon protein predominantly including those amino acids which are common to all naturally-occurring human alpha IFN subtype sequences and including at one or more positions where there is no amino acid common to all subtypes an amino acid which predominantly occurs at that position and in no event including any amino acid residue which is not extant at that position in at least one naturally-occurring subtype.

30 2. Product produced by a microbiological process, comprising a consensus human leukocyte interferon protein according to Claim 1 selected from the group consisting of:

[Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu<sup>158}] IFN- α F;
[Thr¹⁴, Met¹⁶, Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹,
Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu<sup>158}] IFN- α F; and
[Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Glu¹¹⁴, Arg¹²¹, Thr¹⁵⁶, Asn¹⁵⁷, Leu^{158}] IFN- α F.}</sup></sup>

50 3. A gene (DNA molecule) produced by a microbiological process and capable of directing the synthesis in a selected host microorganism of consensus human leukocyte interferon protein according to Claim 1 or 2.

Patentansprüche

55 Patentansprüche für folgende Vertragsstaaten : BE, CH, DE, FR, GB, LI, LU, NL, SE

1. Ein Consensus-Leukozyten-Interferon-Protein, das überwiegend diejenigen Aminosäuren einschließt, die allen natürlich auftretenden Human-alpha-IFN-Subtypsequenzen gemeinsam sind, und das an einer

oder mehreren Positionen, an der es keine Aminosäure gibt, die allen Subtypen gemeinsam ist, eine Aminosäure einschließt, die überwiegend an dieser Position auftritt, und das keinesfalls irgendeinen Aminosäurerest einschließt, der in dieser Position nicht in wenigstens einem natürlich auftretenden Subtyp vorhanden ist.

5 2. Ein Consensus-Human-Leukozyten-Interferon-Protein nach Anspruch 1, ausgewählt aus der Gruppe, bestehend aus:

10 [Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸] IFN- α F;
[Thr¹⁴, Met¹⁶, Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹,
15 Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷,
Leu¹⁵⁸] IFN- α F; und
[Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Glu¹¹⁴, Arg¹²¹, Thr¹⁵⁶, Asn¹⁵⁷,
20 Leu¹⁵⁸] IFN- α F.

25 3. Ein Gen, das in der Lage ist, die Synthese von Consensus-Human-Leukozyten-Interferon-Protein nach Anspruch 1 oder 2 in einem ausgewählten Wirtsmikroorganismus zu steuern.

Patentansprüche für folgenden Vertragsstaat : AT

30 1. Mit einem mikrobiologischen Verfahren hergestelltes Produkt, das ein Consensus-Leukozyten-Interferon-Protein umfaßt, das überwiegend diejenigen Aminosäuren einschließt, die allen natürlich auftretenden Human-alpha-IFN-Subtypsequenzen gemeinsam sind, und das an einer oder mehreren Positionen, an der es keine Aminosäure gibt, die allen Subtypen gemeinsam ist, eine Aminosäure einschließt, die überwiegend an dieser Position auftritt, und das keinesfalls irgendeinen Aminosäurerest einschließt, der in dieser Position nicht in wenigstens einem natürlich auftretenden Subtyp vorhanden ist.

35 2. Mit einem mikrobiologischen Verfahren hergestelltes Produkt, das ein Consensus-Human-Leukozyten-Interferon-Protein umfaßt, nach Anspruch 1, ausgewählt aus der Gruppe, bestehend aus:

40 [Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸] IFN- α F;
[Thr¹⁴, Met¹⁶, Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹,
45 Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷,
Leu¹⁵⁸] IFN- α F; und
[Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Glu¹¹⁴, Arg¹²¹, Thr¹⁵⁶, Asn¹⁵⁷,
50 Leu¹⁵⁸] IFN- α F.

55 3. Ein Gen (DNA-Molekül), das hergestellt ist mit einem mikrobiologischen Verfahren und in der Lage ist, die Synthese von Consensus-Human-Leukozyten-Interferon-Protein nach Anspruch 1 oder 2 in einem

ausgewählten Wirtsmikroorganismus zu steuern.

Revendications

Revendications pour les Etats contractants suivants : BE, CH, DE, FR, GB, LI, LU, NL, SE

5 1. Interféron leucocytaire consensus, incluant, de façon prédominante, les acides aminés qui sont communs à toutes les séquences de sous-types IFN alpha humaines existant naturellement et incluant, en une ou plusieurs positions où il n'y a pas d'acide aminé commun à tous les sous-types, un acide aminé qui existe, de façon prédominante, en cette position et n'incluant en aucun cas un résidu d'acide aminé quelconque qui n'existe pas en cette position dans au moins un sous-type existant naturellement.

10 2. Interféron leucocytaire humain consensus selon la revendication 1, choisi dans le groupe constitué de :

15 [Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Thr¹⁵⁶,
Asn¹⁵⁷, Leu¹⁵⁸]IFN- α F ;

20 [Thr¹⁴, Met¹⁶, Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸]IFN- α F ; et

25 [Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Glu¹¹⁴,
Arg¹²¹, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸]IFN- α F.

30 3. Gène capable de diriger la synthèse dans un micro-organisme hôte choisi d'interféron leucocytaire humain consensus selon la revendication 1 ou 2.

Revendications pour l'Etat contractant suivant : AT

35 1. Produit obtenu par un procédé microbiologique, comprenant un interféron leucocytaire consensus, incluant, de façon prédominante, les acides aminés qui sont communs à toutes les séquences de sous-types IFN alpha humaines existant naturellement et incluant, en une ou plusieurs positions où il n'y a pas d'acide aminé commun à tous les sous-types, un acide aminé qui existe, de façon prédominante, en cette position et n'incluant en aucun cas un résidu d'acide aminé quelconque qui n'existe pas en cette position dans au moins un sous-type existant naturellement.

40 2. Produit obtenu par un procédé microbiologique, comprenant un interféron leucocytaire humain consensus selon la revendication 1, choisi dans le groupe constitué de :

45

50

55

[Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Thr¹⁵⁶,
Asn¹⁵⁷, Leu¹⁵⁸]IFN- α F ;

5

[Thr¹⁴, Met¹⁶, Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰,
Leu⁹⁶, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸]IFN- α F ; et

10

[Arg²², Ala⁷⁶, Asp⁷⁸, Glu⁷⁹, Tyr⁸⁶, Tyr⁹⁰, Leu⁹⁶, Glu¹¹⁴,
Arg¹²¹, Thr¹⁵⁶, Asn¹⁵⁷, Leu¹⁵⁸]IFN- α F.

15

3. Gène (molécule d'ADN) obtenu par un procédé microbiologique et capable de diriger la synthèse dans un micro-organisme hôte choisi d'interféron leucocytaire humain consensus selon la revendication 1 ou 2.

20

25

30

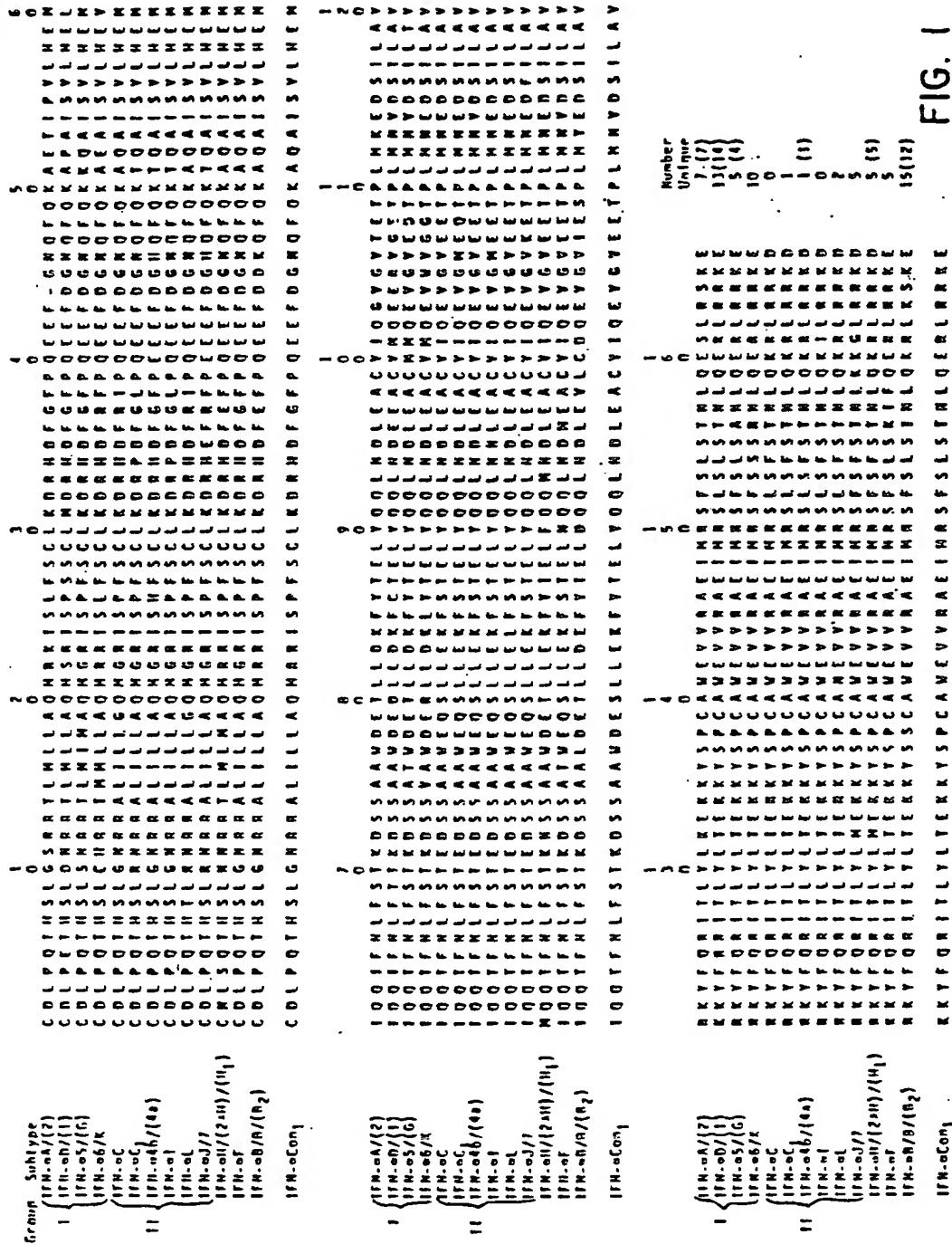
35

40

45

50

55



- 5 -

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.